On-line Supplementary Material

Supplementary Table 1. Hazard ratios and 95% confidence intervals for tertiles of dietary fiber intake and colorectal cancer by the tertiles of folate intake (N=478,994)

Folate intake	Dietary fiber			*P* trend[1]
	Tertile 1 (low)	Tertile 2	Tertile 3 (high)	
Tertile 1 (low)				
Multivariable adjusted[2]	1.00	0.91 (0.84, 0.99)	0.97 (0.84, 1.12)	0.10
Tertile 2				
Multivariable adjusted[2]	1.00	0.95 (0.87, 1.03)	0.93 (0.84, 1.02)	0.15
Tertile 3 (high)				
Multivariable adjusted[2]	1.00	1.02 (0.89, 1.17)	1.07 (0.94, 1.23)	0.17

[1] *P* trend <0.05. All statistical tests were two-sided.

[2] Estimated using Cox proportional hazards regression model adjusted for age (years, continuous), body mass index (<18.5 kg/m^2; 18.5-<25 kg/m^2; 25-<30 kg/m^2; ≥30 kg/m^2, missing), alcohol intake (0 drinks/day; <1 drinks/day; 1-<2 drinks/day; 2-<3 drinks/day; ≥3 drinks/day; missing), general health status (excellent; very good; good; fair; poor; unknown), first degree relatives with colon cancer (yes; no; unknown), race/ethnicity (Non-Hispanic White; Non-Hispanic Black; Hispanic; Asian, Pacific Islander or American Indian/Native American; unknown), education (less than 12 years; 12 years or completed high school; post-high school training other than college; some college; college and post graduate; unknown), sex, physical activity (never, rarely; <3 times/month; 1-2, 3-4, or ≥5 times/week; or missing), smoking (never, ≤20 cigarettes/day in the past, >20 cigarettes/day in the past, ≤20 cigarettes/day currently, >20 cigarettes/day currently, or missing), and intakes of red and processed meat (quintiles), dietary calcium (quintiles), and total energy (kilocalories, continuous). *P* for interaction=0.62.

Supplementary Table 2. Hazard ratios and 95% confidence intervals for tertiles of dietary fiber intake and colorectal cancer by the tertile of red and processed meat (N=478,994)

| Red and processed meat intake | Dietary fiber | | | *P* trend[1] |
	Tertile 1 (low)	Tertile 2	Tertile 3 (high)	
Tertile 1 (low)				
Multivariable adjusted[2]	1.00	0.95 (0.85, 1.07)	1.00 (0.89, 1.13)	0.66
Tertile 2				
Multivariable adjusted[2]	1.00	0.98 (0.89, 1.07)	0.97 (0.87, 1.08)	0.56
Tertile 3 (high)				
Multivariable adjusted[2]	1.00	0.93 (0.86, 1.00)	0.96 (0.86, 1.07)	0.25

[1] *P* trend <0.05. All statistical tests were two-sided.
[2] Estimated using Cox proportional hazards regression model adjusted for age (years, continuous), body mass index (<18.5 kg/m^2; 18.5-<25 kg/m^2; 25-<30 kg/m^2; ≥30 kg/m^2, missing), alcohol intake (0 drinks/day; <1 drinks/day; 1-<2 drinks/day; 2-<3 drinks/day; ≥3 drinks/day; missing), general health status (excellent; very good; good; fair; poor; unknown), first degree relatives with colon cancer (yes; no; unknown), race/ethnicity (Non-Hispanic White; Non-Hispanic Black; Hispanic; Asian, Pacific Islander or American Indian/Native American; unknown), education (less than 12 years; 12 years or completed high school; post-high school training other than college; some college; college and post graduate; unknown), sex, physical activity (never, rarely; <3 times/month; 1-2, 3-4, or ≥5 times/week; or missing), smoking (never, ≤20 cigarettes/day in the past, >20 cigarettes/day in the past, ≤20 cigarettes/day currently, >20 cigarettes/day currently, or missing), and intakes of folate (quintiles), dietary calcium (quintiles), and total energy (kilocalories, continuous). *P* for interaction=0.66.

Supplementary Table 3. Calibration analysis of the association of intake of dietary fiber and whole grains with colorectal cancer incidence using a subset (n = 1,975) with two 24-hour dietary recalls to predict hazard ratios and 95% confidence intervals for intake of dietary fiber and whole grains for the NIH-AARP Diet Health Analytic Cohort (N=478,994)

Fiber variable	Mean (g/1000kcal/day) ± SD	Hazard ratio (HR) for 5 g/1000 kcal/day increase in intake [1]	95% CI [1]	P-value [1,2]
FFQ dietary fiber	10.82 ± 3.91	0.99	(0.95, 1.02)	0.40
Predicted dietary fiber [3]	9.07 ± 2.14	0.97	(0.90, 1.04)	0.40
Whole grains variable	**Mean (servings/1000kcal/day) ± SD**	**Hazard ratio (HR) for 1 serving/1000 kcal/day increase in intake [1]**	**95% CI [1]**	**P-value [1,2]**
FFQ whole grains	0.67 ± 0.45	0.89	(0.85, 0.94)	<0.001
Predicted whole grains [3]	0.50 ± 0.21	0.71	(0.61, 0.82)	<0.001

[1]Estimated using Cox proportional hazards regression model adjusted for age (years, continuous), body mass index (<18.5 kg/m^2; 18.5-<25 kg/m^2; 25-<30 kg/m^2; ≥30 kg/m^2, missing), alcohol intake (0 drinks/day; <1 drinks/day; 1-<2 drinks/day; 2-<3 drinks/day; ≥3 drinks/day; missing), general health status (excellent; very good; good; fair; poor; unknown), first degree relatives with colon cancer (yes; no; unknown), race/ethnicity (Non-Hispanic White; Non-Hispanic Black; Hispanic; Asian, Pacific Islander or American Indian/Native American; unknown), education (less than 12 years; 12 years or completed high school; post-high school training other than college; some college; college and post graduate; unknown), sex, physical activity (never, rarely; <3 times/month; 1-2, 3-4, or ≥5 times/week; or missing), smoking (never, ≤20 cigarettes/day in the past, >20 cigarettes/day in the past, ≤20 cigarettes/day currently, >20 cigarettes/day currently, or missing), and intakes of red and processed meat (linear, squared, and cubic polynomials), dietary calcium (linear, squared, and cubic polynomials), dietary folate (linear, squared, and cubic terms), and total energy (kilocalories, continuous).

[2] *P* trend <0.05. All statistical tests were two-sided.

[3] Predicted intake estimated using multivariable linear regression model with mean nutrient-density adjusted dietary fiber or whole grains intake based on two days of 24-hour dietary recalls as the response variable and mean FFQ nutrient-density adjusted dietary fiber or whole grains intake as the main predictor and adjusted for age (years, continuous), body mass index (<18.5 kg/m^2; 18.5-<25 kg/m^2; 25-<30 kg/m^2; ≥30 kg/m^2, missing), alcohol intake (0 drinks/day; <1 drinks/day; 1-<2 drinks/day; 2-<3 drinks/day; ≥3 drinks/day; missing), general health status (excellent; very good; good; fair; poor; unknown), first degree relatives with colon cancer (yes; no; unknown), race/ethnicity (Non-Hispanic White; Non-Hispanic Black; Hispanic; Asian, Pacific Islander or American Indian/Native American; unknown), education (less than 12 years; 12 years or completed high school; post-high school training other than college; some college; college and post graduate; unknown), sex, physical activity (never, rarely; <3 times/month; 1-2, 3-4, or ≥5 times/week; or missing), smoking (never, ≤20 cigarettes/day in the past, >20 cigarettes/day in the past, ≤20 cigarettes/day currently, >20 cigarettes/day currently, or missing), and intakes of red and processed meat (quintiles), dietary calcium (quintiles), folate (quintiles) and total energy (kilocalories, continuous).

Supplementary Table 4. Hazard ratios and 95% confidence intervals for the residuals of whole grains and dietary fiber intake and colorectal cancer (N=478,994)

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	*P* trend[1]
Whole grains						
Multivariable adjusted[2]	1.00	0.97 (0.91, 1.03)	0.93 (0.88, 0.99)	0.88 (0.82, 0.93)	0.88 (0.83, 0.93)	<0.0001
Dietary fiber						
Multivariable adjusted[2]	1.00	0.99 (0.93, 1.05)	0.98 (0.92, 1.04)	1.01 (0.95, 1.08)	0.97 (0.91, 1.03)	0.57

[1] *P* trend <0.05. All statistical tests were two-sided.

[2] Estimated using Cox proportional hazards regression model adjusted for age (years, continuous), body mass index (<18.5 kg/m^2; 18.5-<25 kg/m^2; 25-<30 kg/m^2; ≥30 kg/m^2, missing), alcohol intake (0 drinks/day; <1 drinks/day; 1-<2 drinks/day; 2-<3 drinks/day; ≥3 drinks/day; missing), general health status (excellent; very good; good; fair; poor; unknown), first degree relatives with colon cancer (yes; no; unknown), race/ethnicity (Non-Hispanic White; Non-Hispanic Black; Hispanic; Asian, Pacific Islander or American Indian/Native American; unknown), education (less than 12 years; 12 years or completed high school; post-high school training other than college; some college; college and post graduate; unknown), sex, physical activity (never, rarely; <3 times/month; 1-2, 3-4, or ≥5 times/week; or missing), smoking (never, ≤20 cigarettes/day in the past, >20 cigarettes/day in the past, ≤20 cigarettes/day currently, >20 cigarettes/day currently, or missing), and intakes of red and processed meat (linear, squared, and cubic polynomials), dietary calcium (linear, squared, and cubic polynomials), dietary folate (linear, squared, and cubic terms), and total energy (kilocalories, continuous).

Supplementary Table 5. Hazard ratios and 95% confidence intervals for quintiles dietary fiber intake, by type and source, and colorectal cancer subsite (N=478,994)

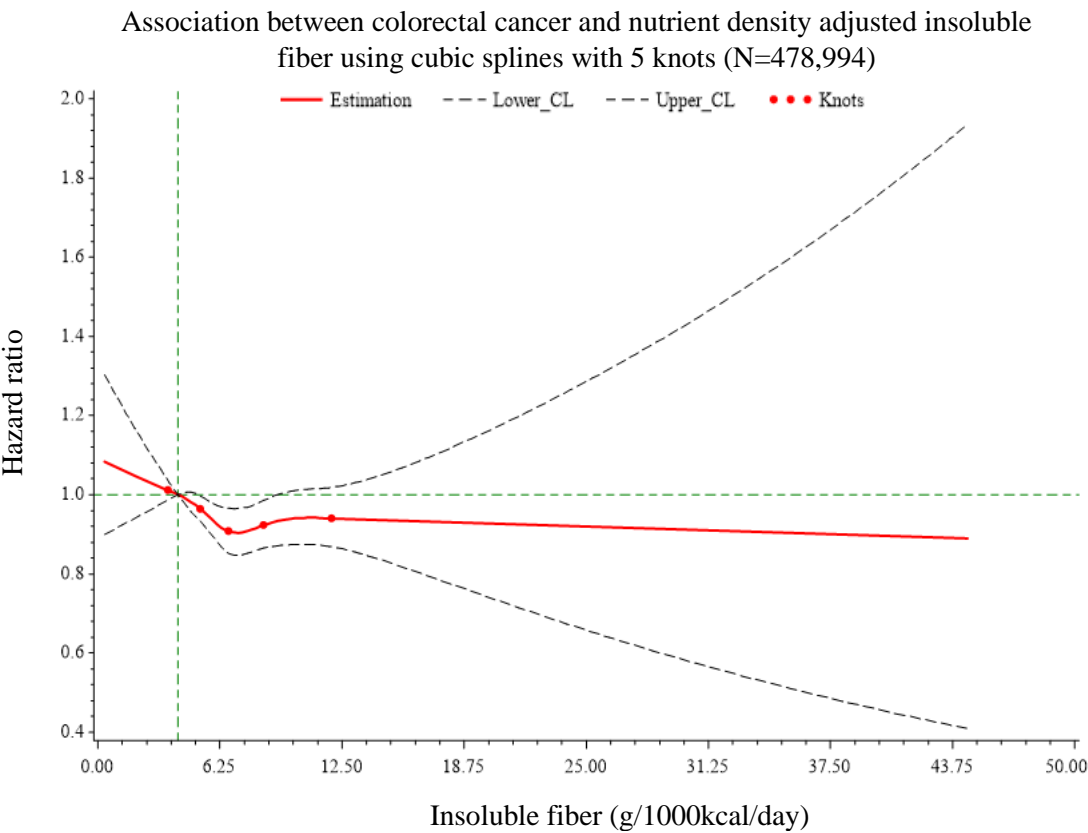|	Dietary fiber					
Soluble fiber	**Quintile 1**	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5**	***P* trend[1]**
Colon						
Multivariable adjusted[2]	1.00	0.95 (0.88, 1.02)	0.91 (0.85, 0.99)	0.92 (0.85, 1.00)	0.99 (0.90, 1.08)	0.89
Proximal colon						
Multivariable adjusted[2]	1.00	0.91 (0.82, 1.00)	0.91 (0.82, 1.01)	0.88 (0.79, 0.99)	0.99 (0.88, 1.11)	0.87
Distal colon						
Multivariable adjusted[2]	1.00	0.99 (0.89, 1.11)	0.87 (0.77, 0.99)	0.95 (0.83, 1.08)	0.97 (0.84, 1.12)	0.62
Rectum						
Multivariable adjusted[2]	1.00	1.07 (0.95, 1.21)	1.09 (0.96, 1.24)	1.17 (1.02, 1.35)	1.13 (0.97, 1.32)	0.08
Insoluble fiber	**Quintile 1**	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5**	***P* trend[1]**
Colon						
Multivariable adjusted[2]	1.00	0.96 (0.89, 1.03)	0.96 (0.89, 1.04)	0.88 (0.81, 0.96)	0.97 (0.88, 1.07)	0.36
Proximal colon						
Multivariable adjusted[2]	1.00	0.95 (0.86, 1.05)	0.96 (0.87, 1.07)	0.90 (0.81, 1.01)	0.97 (0.86, 1.09)	0.60
Distal colon						
Multivariable adjusted[2]	1.00	0.97 (0.86, 1.08)	0.96 (0.849, 1.09)	0.83 (0.72, 0.95)	0.95 (0.82, 1.11)	0.28
Rectum						
Multivariable adjusted[2]	1.00	0.95 (0.84, 1.07)	0.96 (0.84, 1.09)	0.96 (0.83, 1.10)	0.99 (0.84, 1.15)	0.97
Fruit fiber	**Quintile 1**	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5**	***P* trend[1]**
Colon						
Multivariable adjusted[2]	1.00	0.88 (0.82, 0.94)	0.91 (0.84, 0.98)	0.89 (0.82, 0.96)	0.95 (0.87, 1.03)	0.63
Proximal colon						
Multivariable adjusted[2]	1.00	0.89 (0.81, 0.98)	0.88 (0.80, 0.97)	0.91 (0.82, 1.01)	0.95 (0.86, 1.06)	0.94
Distal colon						
Multivariable adjusted[2]	1.00	0.85 (0.76, 0.96)	0.96 (0.85, 1.08)	0.86 (0.76, 0.98)	0.94 (0.82, 1.07)	0.55
Rectum						
Multivariable adjusted[2]	1.00	1.02 (0.90, 1.14)	0.96 (0.85, 1.09)	1.11 (0.97, 1.26)	1.07 (0.93, 1.23)	0.17
Vegetable fiber	**Quintile 1**	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5**	***P* trend[1]**
Colon						
Multivariable adjusted[2]	1.00	0.93 (0.86, 1.00)	0.88 (0.82, 0.95)	0.93 (0.86, 1.00)	0.98 (0.91, 1.06)	0.95
Proximal colon						
Multivariable adjusted[2]	1.00	0.94 (0.86, 1.03)	0.90 (0.82, 0.99)	0.94 (0.85, 1.03)	0.99 (0.89, 1.10)	0.86
Distal colon						
Multivariable adjusted[2]	1.00	0.92 (0.82, 1.03)	0.87 (0.77, 0.98)	0.94 (0.83, 1.06)	0.96 (0.84, 1.09)	0.82
Rectum						
Multivariable adjusted[2]	1.00	0.92 (0.81, 1.03)	1.01 (0.89, 1.14)	0.96 (0.85, 1.09)	1.00 (0.88, 1.15)	0.73
Bean fiber	**Quintile 1**	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5**	***P* trend[1]**
Colon						
Multivariable adjusted[2]	1.00	0.99 (0.92, 1.06)	1.00 (0.93, 1.70)	0.95 (0.89, 1.03)	0.97 (0.90, 1.05)	0.36
Proximal colon						
Multivariable adjusted[2]	1.00	0.96 (0.88, 1.06)	1.03 (0.94, 1.14)	1.01 (0.92, 1.12)	1.02 (0.93, 1.12)	0.46
Distal colon						
Multivariable adjusted[2]	1.00	1.01 (0.90, 1.13)	0.96 (0.85, 1.08)	0.91 (0.81, 1.02)	0.94 (0.83, 1.06)	0.16
Rectum						
Multivariable adjusted[2]	1.00	0.99 (0.88, 1.12)	0.94 (0.83, 1.07)	1.07 (0.95, 1.21)	1.05 (0.93, 1.19)	0.15
Grain fiber	**Quintile 1**	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5**	***P* trend[1]**

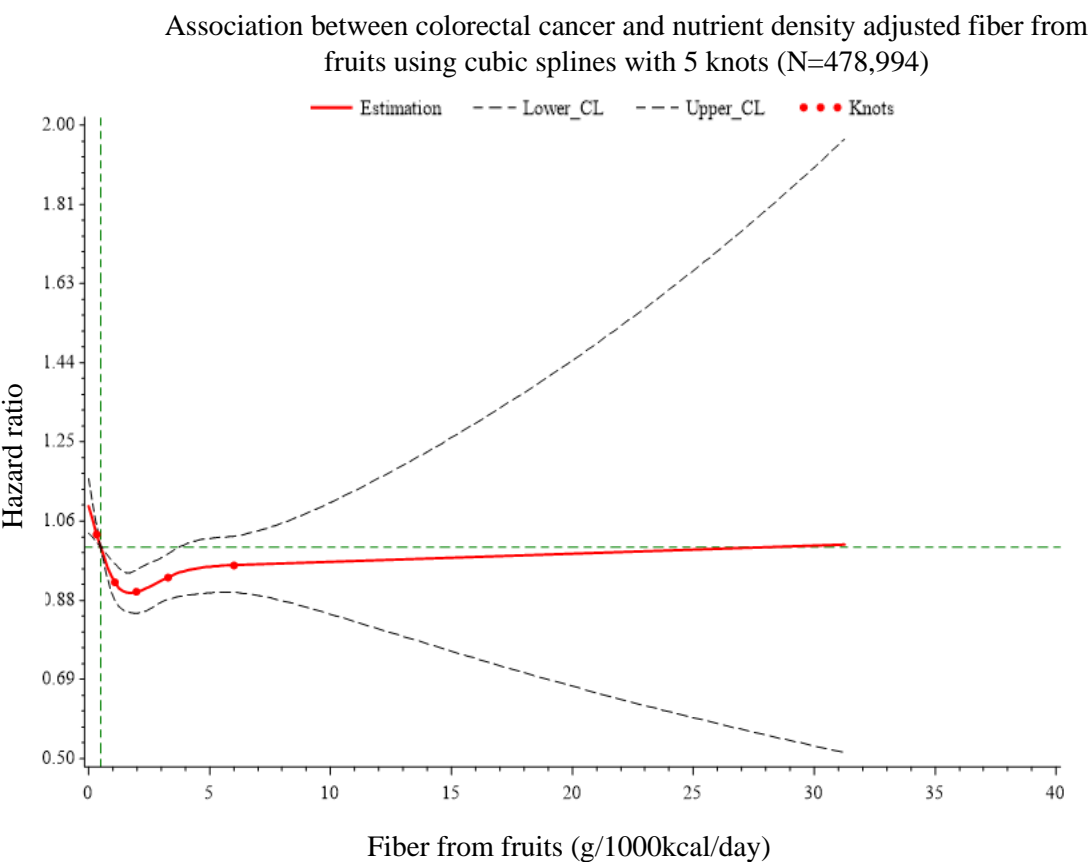|					
Colon					
Multivariable adjusted[2]	1.00	1.01 (0.94, 1.09)	0.99 (0.92, 1.07)	0.97 (0.90, 1.05)	0.94 (0.87, 1.02)	0.08
Proximal colon					
Multivariable adjusted[2]	1.00	1.03 (0.94, 1.13)	1.02 (0.92, 1.13)	1.04 (0.94 1.15)	1.00 (0.894, 1.12)	0.88
Distal colon					
Multivariable adjusted[2]	1.00	0.97 (0.87, 1.09)	0.97 (0.87, 1.10)	0.88 (0.78, 1.00)	0.84 (0.73, 0.96)	0.01
Rectum					
Multivariable adjusted[2]	1.00	0.93 (0.83, 1.04)	0.88 (0.78, 1.00)	0.84 (0.74, 0.96)	0.77 (0.66, 0.88)	<0.001

[1] *P* trend <0.05. All statistical tests were two-sided.

[2] Estimated using Cox proportional hazards regression model adjusted for age (years, continuous), body mass index (<18.5 kg/m^2; 18.5-<25 kg/m^2; 25-<30 kg/m^2; ≥30 kg/m^2, missing), alcohol intake (0 drinks/day; <1 drinks/day; 1-<2 drinks/day; 2-<3 drinks/day; ≥3 drinks/day; missing), general health status (excellent; very good; good; fair; poor; unknown), first degree relatives with colon cancer (yes; no; unknown), race/ethnicity (Non-Hispanic White; Non-Hispanic Black; Hispanic; Asian, Pacific Islander or American Indian/Native American; unknown), education (less than 12 years; 12 years or completed high school; post-high school training other than college; some college; college and post graduate; unknown), sex, physical activity (never, rarely; <3 times/month; 1-2, 3-4, or ≥5 times/week; or missing smoking (never, ≤20 cigarettes/day in the past, >20 cigarettes/day in the past, ≤20 cigarettes/day currently, >20 cigarettes/day currently, or missing), and intakes of folate (quintiles), intakes of red and processed meat (quintiles), dietary calcium (quintiles), and total energy (kilocalories, continuous).

Supplementary Figure 1.

A

Association between colorectal cancer and nutrient density adjusted insoluble fiber using cubic splines with 5 knots (N=478,994)

Insoluble fiber (g/1000kcal/day)

B

Association between colorectal cancer and nutrient density adjusted fiber from fruits using cubic splines with 5 knots (N=478,994)
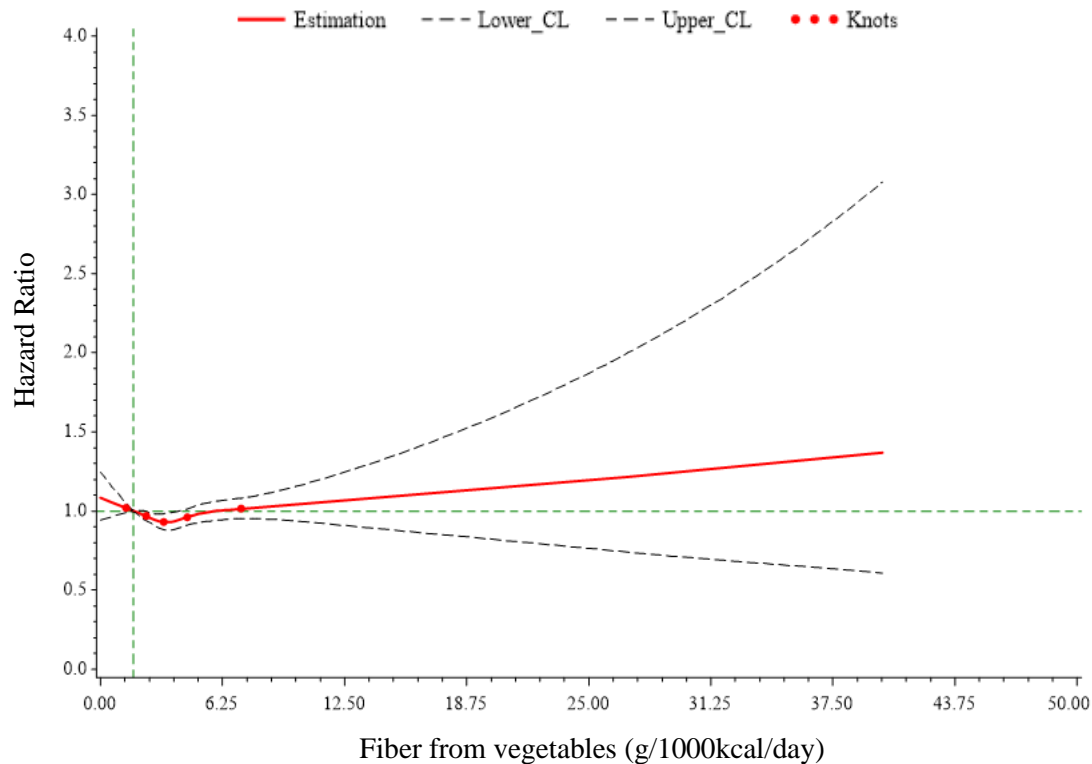

Fiber from fruits (g/1000kcal/day)

C

Association between colorectal cancer and nutrient density adjusted fiber from
vegetables using cubic splines with 5 knots (N=478,994)

Supplementary Figure 1. Nonlinear associations for A) nutrient density adjusted insoluble fiber B)
nutrient density adjusted fiber from fruits and C) nutrient density adjusted fiber from vegetables and
colorectal cancer estimated using Cox proportional hazards regression models and restricted cubic splines
for dietary fiber type (i.e., insoluble) or source (i.e., fruits and vegetables) where the reference value for
dietary fiber type or source was set at the median intake value of the first quartile (vertical green dashed-
line) for HR estimates with 5 knots set at the 5th, 25th, 50th, 75th, and 95th percentiles of dietary fiber
intake (in red) and 95% confidence intervals (black dashed lines). P-values for nonlinear associations
between dietary fiber type or source and colorectal cancer risk were estimated using a likelihood ratio
test. All models were adjusted for age (years, continuous), body mass index (<18.5 kg/m^2; 18.5-<25
kg/m^2; 25-<30 kg/m^2; ≥30 kg/m^2, missing), alcohol intake (0 drinks/day; <1 drinks/day; 1-<2 drinks/day;
2-<3 drinks/day; ≥3 drinks/day; missing), general health status (excellent; very good; good; fair; poor;
unknown), first degree relatives with colon cancer (yes; no; unknown), race/ethnicity (Non-Hispanic
White; Non-Hispanic Black; Hispanic; Asian, Pacific Islander or American Indian/Native American;
unknown), education (less than 12 years; 12 years or completed high school; post-high school training
other than college; some college; college and post graduate; unknown), sex, physical activity (never,
rarely; <3 times/month; 1-2, 3-4, or ≥5 times/week; or missing smoking (never, ≤20 cigarettes/day in the
past, >20 cigarettes/day in the past, ≤20 cigarettes/day currently, >20 cigarettes/day currently, or
missing), and intakes of folate (quintiles), intakes of red and processed meat (quintiles), dietary calcium
(quintiles), and total energy (kilocalories, continuous).

On-line Supplementary Methods

Dietary fiber and whole grains sub-study using 24-hour dietary recalls

The analytic cohort included n=1,975 participants who completed two 24-hour dietary recalls on non-consecutive days. Dietary fiber was estimated in grams per day and whole grains intake was estimated in servings per day. Values were averaged for the two days of recall. A multivariable linear regression model predicting two-day mean dietary fiber intake and whole grains intake in this subset was fit using continuous variables for dietary fiber and whole grains, as well as the *a priori* selected potential confounders (i.e., age (years, continuous), body mass index (<18.5 kg/m^2; 18.5-<25 kg/m^2; 25-<30 kg/m^2; ≥30 kg/m^2, missing), alcohol intake (0 drinks/day; <1 drinks/day; 1-<2 drinks/day; 2-<3 drinks/day; ≥3 drinks/day; missing), general health status (excellent; very good; good; fair; poor; unknown), first degree relatives with colon cancer (yes; no; unknown), race/ethnicity (Non-Hispanic White; Non-Hispanic Black; Hispanic; Asian, Pacific Islander or American Indian/Native American; unknown), education (less than 12 years; 12 years or completed high school; post-high school training other than college; some college; college and post graduate; unknown), sex, physical activity (never, rarely; <3 times/month; 1-2, 3-4, or ≥5 times/week; or missing), smoking (never, ≤20 cigarettes/day in the past, >20 cigarettes/day in the past, ≤20 cigarettes/day currently, >20 cigarettes/day currently, or missing), and intakes of red and processed meat (quintiles), dietary calcium (quintiles), and total energy (kilocalories, continuous)). The intercept and beta coefficients from the linear regression model were subsequently used to predict the mean two-day dietary fiber and whole grains intake for each participant in the analytic cohort (n=478,994). Predicted mean two-day dietary fiber intake was standardized by dividing the gram value by a factor of 5. The scaled predicted value was then modeled as a continuous variable in the multivariable-adjusted Cox proportional hazards regression model to estimate the hazard ratio and confidence interval for a 5-gram per day increase in dietary fiber intake and a one serving increase in whole grains intake with colorectal cancer with age as the underlying time metric.

Residual method of adjustment for dietary variables

Because dietary variables were highly correlated, we used the residual method to adjust for these variables. First, a linear regression model predicting whole grains (as quintiles) was fit using continuous dietary variables as linear, squared, and cubic polynomials including red meat, folate, and calcium. Residuals of whole grains were obtained and made into quintiles. The quintiles of the whole grains residuals were then used in a multivariable-adjusted Cox proportional hazards regression model with adjustment for the dietary variables as linear, squared, and cubic polynomials as well as the *a priori* selected potential confounders (i.e., age (years, continuous), body mass index (<18.5 kg/m^2; 18.5-<25 kg/m^2; 25-<30 kg/m^2; ≥30 kg/m^2, missing), alcohol intake (0 drinks/day; <1 drinks/day; 1-<2 drinks/day; 2-<3 drinks/day; ≥3 drinks/day; missing), general health status (excellent; very good; good; fair; poor; unknown), first degree relatives with colon cancer (yes; no; unknown), race/ethnicity (Non-Hispanic White; Non-Hispanic Black; Hispanic; Asian, Pacific Islander or American Indian/Native American; unknown), education (less than 12 years; 12 years or completed high school; post-high school training other than college; some college; college and post graduate; unknown), sex, physical activity (never, rarely; <3 times/month; 1-2, 3-4, or ≥5 times/week; or missing), smoking (never, ≤20 cigarettes/day in the past, >20 cigarettes/day in the past, ≤20 cigarettes/day currently, >20 cigarettes/day currently, or missing), and intakes of red and processed meat (quintiles), dietary calcium (quintiles), and total energy (kilocalories, continuous)). The same method was applied to dietary fiber as the dependent variable.